DOCUMENT RESUME

ED 177 150 SP 014 899

AUTHOR	Ward, Beatrice A.
TITLE	Assessment of Teacher Performance: What Is Involved? What Is the Cost?
INSTITUTION	Far West Lab. for Educational Research and Development, San Francisco, Calif.
PUB DATE	Feb 73
NOTE	22p.; Paper presented at the Multi-State Consortium on Performance Based Teacher Education (New Orleans, Louisiana, February 27, 1973)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Academic Achievement; *Cost Effectiveness; *Educational Assessment; *Evaluation Methods; Performance Based Teacher Education; *Program Budgeting; Teacher Behavior; *Teacher Evaluation

ABSTRACT

Questions are raised regarding the purposes of teacher assessment, the teacher performance to be assessed, and the level of training and skill individuals who do the assessing must have. In each question area, several courses of action are presented. Different cost levels are associated with each option. It is suggested that the task for the individual, or agency, responsible for assessing teacher performance is to determine which options within each decision area best meet the policy-making and training needs of that particular program. Central to this decision is the quantity and quality of data required given the purposes for which the findings will be used. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED177150

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ASSESSMENT OF TEACHER PERFORMANCE: WHAT IS INVOLVED? WHAT IS THE COST?*

Beatrice A. Ward
Far West Laboratory for Educational Research and Development

When the term "performance-based" is used as a descriptor for teacher training and/or teacher certification, certain essential conditions apply. What the teacher does is as important as what he knows. During training, skill development receives equal or greater attention than knowledge acquisition. At the end of training, mastery is used as the evaluative criteria rather than a list of courses completed and grades received. Information about the teacher's performance is obtained throughout the training period and subsequently in the operable classroom in order to determine whether the specified teaching skills are being acquired and used. Student performance also is studied in order to determine whether teacher use of certain skills relates to the level and quality of student learning.

Within such a framework, assessment of teacher performance is multi-faceted. It is conducted for many different reasons, focuses upon a wide variety of teacher and student behaviors, and involves a diversity of individuals and institutions. Inquiry into several of these facets of assessment is the purpose of this paper. The discussion centers around choices that may be made regarding assessment and the costs associated with particular options. It builds upon the information obtained by the

---

* Invited paper, Multi-State Consortium on Performance-Based Teacher Education, New Orleans, Louisiana, February 27, 1973.

SP 014 899

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Far West Laboratory for Educational Research and Development
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Teacher Education Division at the Far West Laboratory for Educational Research and Development during five years of developing and testing performance-based teacher training materials.

Why Assess Teacher Performance

Assessment of teacher performance can be undertaken for a variety of purposes. Among them are:

- To determine the average type of performance that can be expected of experienced or inexperienced teachers at a local, regional, state, or national level.
- To identify the types of teaching skills a majority of inservice or preservice teachers perform well and those which they need to improve or acquire.
- To establish a minimum acceptable level of performance for success as a teacher; to identify those teaching skills that relate to successful student performance.
- To determine what an individual's level of performance is at the time he enters inservice or preservice teacher training.
- To establish that an individual has mastered one or more teaching skills as a result of training.

Two types of data are associated with these forms of assessment. Assessment conducted to determine average teacher performance, identify how well teachers perform certain specific skills, or establish a minimum acceptable level of performance builds upon information regarding teaching in general. Samples of teaching and teachers may be used as a data base. The information obtained serves as a guide for policy decisions such as the selection of the content (skills and knowledge) to be included in a pre-service and/or inservice training program and the specification of the skill and knowledge requirements for certification. General information of this type also aids in determining how extensive an inservice retraining program should be; i.e., the potential number of teachers needing training

in a given skill and the georgaphic areas in which particular types of training are needed.

On the other hand, the data base for planning an individual teacher's training program and/or verifying that a particular individual has acquired certain specified teaching skills requires detailed information regarding that person's performance. Sampling procedures cannot be used. Rather, a profile of skill usage along with a judgment as to the quality of use must be available for each individual teacher. The specificity of the profile may vary depending upon whether the pattern of skill use will be used to outline a training sequence or verify competence prior to certification. Pupil learning outcomes may or may not be part of the profile depending upon the inservice or preservice status of the teacher. But, regardless of the information included, the assessment process must help each individual teacher identify his areas of strength and weakness and plan a training program that will improve his teaching performance.

Assessment of the individual teacher or assessment of a sample of the teacher population, therefore, evolves from the intended use of the assessment information. Policy decisions can build from a different data base than training decisions. Obviously, given sufficient time and financial resources, assessment of individual teachers could serve policy-making as well as training purposes. However, since time and money are limited, study of a representative sample of teachers appears to be the most cost/effective means for obtaining the data needed to make policy decisions. A variety of techniques may be used to arrive at the teaching sample. For example, an adaptation of matrix sampling (Husek and Sirotnik, 1968; Lord and Novick, 1968) might provide a means for obtaining information about how a broad sample of teachers' use of a wide-range of skills

at the same time keeping costs similar to those associated with an indepth study of a small sample.

Based upon these guidelines, the ultimate cost of teacher assessment is determined, in part, by the purposes for which the findings will be used. A first step toward designing and financing an assessment effort is to outline the questions the data are to answer.

What Teaching Performance Will Be Assessed?

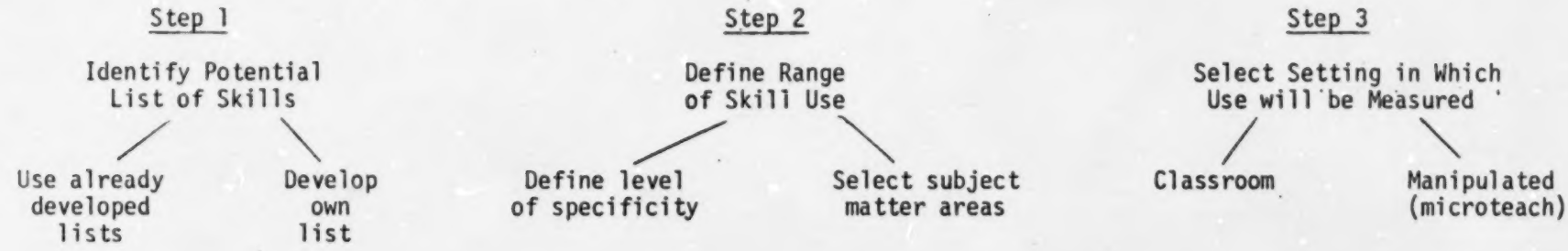
Determining what teaching performance will be assessed also involved several steps (see Figure 1). The first is to identify the potential list of skills to be included in the assessment effort.

Identifying skills. The diversity of knowledge and skills encompassed within the teaching act has been documented by numerous experts in the field: i.e., Smith, 1971; Houston, 1972; Joyce & Weil, 1972. Yet, to date, research has shown few positive relationships between teaching skill and student outcomes (Rosenshine, 1971). As a result, selection of the skills to be assessed, for the most part, consists of combining what little evidence exists regarding teacher effects with theoretical and practical opinion regarding desired teacher performance.

Several efforts to identify critical teaching skills have already been undertaken by individuals and agencies such as the developers of the models for elementary teacher education (Elementary Teacher Education Models, 1969), the designers of competency-based teacher training programs (i.e., Dodl, 1972), and the developers of performance-based teacher training materials (i.e., the Far West Laboratory). In each instance, a thorough review of the literature related to a particular aspect of teaching is undertaken, in-class observations are made to verify that certain skills are used, and experienced teachers, school

Figure 1

Steps in Selecting Teaching Performance to Be Assessed



administrators, parents, and others are asked to judge the relative importance of the skills.

As suggested by Joyce (1973), persons responsible for teacher assessment may opt to use these already developed skill lists or they may choose to develop their own. If the latter option is taken, the costs presented in Table 1 indicate the expenditures in personnel and other operating expenses such as travel and purchase of supplies and materials that may be required. The figures are based upon costs incurred by the Teacher Education Division of the Far West Laboratory while identifying critical teaching skills in the areas of pupils' language development, mathematics tutoring, independent learning, and higher cognitive questioning. On the average, skill identification cost $14,670 for each set of skills.

Table 1

Costs of Identifying a Set of Teaching Skills

Skill Area

	Pupils' Language Development	Math Tutoring	Independent Learning	Higher Cognitive Questioning	Mean
Personnel Costs	$ 5,584	$10,068	$ 4,953	$ 5,981	$ 6,647
Direct Costs	6,938	11,836	6,156	7,163	8,023
TOTAL	$12,522	$21,904	$11,109	$13,144	$14,670

The efficacy of using and/or building upon existing skill lists is apparent given these costs. Since teaching incorporates multiple sets of

skills of the type identified by the Laboratory, the total cost of skill identification could be expected to exceed $100,000 if started from the beginning. This would be the cost even if the search were to be limited to available research information, and theoretical and experiential opinion. If further research to identify relationships between teacher and pupil performance were included, the costs would be much higher. Thus, if assessment costs are to be kept within reason, the skills to be assessed should be taken from the skill lists being compiled as part of the broad performance-based teacher education effort. Each assessment agency should not attempt to develop its own unique skills list.

Range of skill use. The second step in determining what teacher performance will be assessed is to define the range of skill use to be studied. By this, I mean the level of specificity at which the skill will be assessed and the variety of subject matter areas in which the skill samples will be taken.

Many areas of teaching can be described in terms of general performance levels as well as specific categories within these levels. For example, the general teaching skill of asking higher cognitive questions can be divided into the more specific skills of asking analysis, synthesis, and evaluation questions (see Table 2). Within any assessment process, whether it is for policy-making or training purposes, measures of this skill probably would focus upon the general performance level. Previous research indicates that higher cognitive questions generally represent less than half the questions asked by teachers (Gallagher, 1965; Davis & Tinsley, 1967; Guszak, 1967). In order to set training priorities and/or verify competence in this area, therefore, the single measure of percentage of higher cognitive questions asked is adequate. If a teacher

Table 2

Higher Cognitive Questioning*

General Skill Level: Use higher cognitive questions

Specific Skill Level:

1. Use three types of analysis questions.
   - questions requiring students to think of motives or causes
   - questions requiring students to make inferences
   - questions requiring students to find evidence to support generalizations, interpretations, or conclusions
2. Use two types of synthesis questions
   - questions asking students to make predictions
   - questions asking students to develop solutions to problems
3. Use three types of evaluation questions.
   - questions asking students to take a stand on a controversial issue
   - questions asking students to judge truth or validity
   - questions asking students to judge beauty or worth

* Taken from Minicourse 9, Higher Cognitive Questioning, developed by Far West Laboratory.

(preservice or inservice) asks more than 60-70 percent higher cognitive questions, he is performing much better than the average teacher. If he is using considerably less than 50 percent higher cognitive questions, he probably would benefit from training. Given this general piece of assessment information, many decisions can be made. It is only after a decision to design and instigate training has been made that the specific skill levels become important. Learning to ask analysis, synthesis, and evaluation questions can help increase a teacher's over-all use of higher cognitive questions. Specificity at the training stage is essential. For assessment purposes, general measures of skill use can provide adequate information and at the same time keep measurement efforts and costs within reasonable limits.

As outlined in Figure 1, defining the range of skill use to be assessed also includes selection of the subject matter areas in which teacher performance will be studied. Again, this is a decision that affects assessment costs. Single examples of teaching are less costly to obtain than multiple examples.

For some teaching skills, the decision is obvious. Skills related to teaching reading should be assessed in the context of a reading lesson. Likewise skills specific to the teaching of mathematics need to be measured while mathematics is being taught.

Other skills are more general in nature. Probing, for example, can be used in any subject area. It employs the same type of teacher-pupil interaction regardless of the content of the lesson. A teacher's use of probing in any subject area probably would be representative of his use of the skill in general.

Still other skills differ in use depending upon the context in which they are applied; for example, response to pupils' errors in reading probably differs from response to errors in science.

The problem within an assessment effort is to determine which subject areas are most likely to offer the best opportunity for teachers to exhibit their use of each particular skill. During the initial phases of assessment, inquiry probably will be limited to generalizable skills or to specific content area skills. Because of data collection complexity, information regarding skills that apply to several content areas but differ in their use within each area will follow later.

Performance setting. The third step in selecting the teaching performance to be assessed is to select the setting in which skill use will be measured. Based upon Turner's (1972) six criteria of teacher performance, in-classroom as contrasted with manipulated examples of behavior may be considered; the most common manipulated example being a microteach lesson. The salient question to be considered is whether a reasonable sample of skill use can be obtained in a microteach setting.

Inasmuch as much of teaching occurs in small group or one-to-one tutorial situations and critics of the classroom (such as Silberman, 1970) contend that large group instruction should seldom take place, a small group setting, e.g., a microteach lesson, may be an appropriate information base for studying a high proportion of teaching skills. The disadvantage of such a performance sample is in regard to student outcomes. Few student outcomes of any consequence could be expected to occur in a single 10-20 minute lesson. Thus, if study of teacher effect upon student learning is to be included in the assessment process, at least some examples of in-class performance should be included. It

should be noted, however, that given the present state of the art of assessment, researchers (Smith, 1971; Herbert, 1971) question the appropriateness of using student outcomes to measure teacher performance. Further, since the restricted environment of a microteach lesson, whether live or on videotape, also facilitates concurrent scoring of a number of teaching skills, this instructional setting warrants serious consideration as an assessment tool.

Cost of assessment. The Teacher Education Division of the Far West Laboratory has conducted a number of studies of teacher change resulting from Minicourse training. In these studies, an assessment procedure has been used that includes:

- assessment of already identified skills (skills were identified during development of training materials),
- assessment of both general and specific skill use (for example, both percent of higher cognitive questions and use of analysis, synthesis, and evaluation questions were assessed),
- assessment within a certain prescribed content area,
- assessment within a microteach setting.

Using this procedure, microteach lessons were recorded on videotape and critiqued at some later date by trained observers. Only limited, if any, measures of student performance were obtained.

Sample costs for conducting the assessment are presented in Table 3. They are based on the study of five teachers because this is the number of teachers who can conveniently be scheduled into a microteach facility during a 9:00 a.m. to 3:30 p.m. school day. The average cost per teacher for 5 teachers is $180. Since critiquer training, research design, and

Table 3

Cost of Assessing Teacher Performance
Using a Microteach Setting*

Collecting Performance Sample		
5 1/2 hour videotapes @ $15/each	$ 75.00	
7.5 hours of research intern to monitor collection of videotape examples (@ $9450/year + benefits)	50.00	
		$ 120.00
Critiquing Performance		
Critiquer training; one hour training per behavior		
12 behaviors x 2 critiquers x 1 hour		
24 hours x $3.50/hour	84.00	
12 hours of research intern to do training	77.00	
Critiquers to score recorded lessons		
12 behaviors = 6 passes to score		
6 passes @ 30 min. each = 3 hours per tape		
3 hours x 5 tapes = 15 hours @ $3.50 per hour		
15 hours @ $3.50 per hour x 2 critiquers	105.00	
		266.00
Research design and data interpretation		
One week research design planning @ $17,000/year (Note: If more teachers were involved, this cost would be dispersed over the entire population)	434.00	434.00
Data preparation		
keypunch - 1 day clerk/analyst at $9,000/year + benefits	40.00	
Run and analyze - 1 day	40.00	
		80.00
TOTAL COST (Assessment of 5 teachers)		$900.00

* Based on assessment of 5 teachers

data preparation costs would not be repeated with a larger sample of teachers, the cost to assess 30 teachers would be $2160; or $72 per teacher.

If the skills to be measured and the lesson content were carefully selected, such an assessment effort could provide a large amount of information about teacher performance. It could be used for many of the purposes outlined earlier in this paper.

Who Will Do the Assessing?

Assessment of teacher performance generally will be conducted by three types of individuals.

1. Teachers. Peer observations (one teacher observing another and vice versa) can be used to study a wide variety of teaching skills.
2. Critiquers. Analysis of audio and/or videotape recordings of teaching (e.g., a microteach lesson) may be done by specially trained critiquers. They also may be teachers but more often they will be graduate or undergraduate students in education hired to carry out a particular critiquing task.
3. In-class observers. Whenever an assessment program demands high accuracy in recording teacher in-class performance, specially trained individuals will be needed. In-class scoring of teacher performance requires simultaneous monitoring of multiple variables and instantaneous recognition of the skill(s) to be assessed. Considerable training and practice must be completed in order to achieve this degree of observer skill.

Information obtained through peer observations will be the least accurate form of assessment data. Nonetheless, when the assessment is being done to identify inservice teachers' training needs, this type of information is sufficient to establish a tentative skill profile.

In our work at the Far West Laboratory, we tested a form of peer assessment in the responsive skills area. Six observations were made (in this case, four were done by teacher peers and two were completed by the teacher's students). Gross information was obtained regarding such skills as teacher response to pupil ideas, teacher use of positive and/or negative responses to student performance, and teacher use of nonverbal reinforcement. Given the observation findings, the teachers were alerted to skill areas in which they needed to improve. Decisions could be made regarding where to begin the inservice training program for that group of teachers. The gross information received from the peer observations allowed the training to proceed with some degree of individual focus.

Critiquer analysis of audio or videotape recordings of teaching samples provides more precise information about the skills used. The data are more reliable than peer observations. Many skills can be scored from a single teaching sample by rerunning the recording several times. On the other hand, since the teaching situation being studied is manipulated, questions must be asked regarding the generalizability of these findings to the teacher's regular performance in the classroom. The best time to use this form of assessment may be at the end of training when a teacher needs to be sure that he has acquired and is using a particular skill or set of skills and when nuances in skill use need to be pursued.

Overall, the most comprehensive observation procedure is to have highly skilled observers score teacher and pupil performance in the classroom setting. As noted above, this requires the observer to identify specific skills at the moment they occur. Teacher-pupil interaction in an on-going classroom setting cannot be rerun to check the observer's scoring. Further, since the observer can code the use of only a small number of skills at a time, several observations may be necessary in order to code the same number of skills that could be scored in a single 30-minute videotaped lesson. In-class observations appear to be most useful when pupil outcomes are to be assessed as well as teacher performance, when a particular teaching skill is expected to occur only in a natural setting, when data on teacher performance must be analyzed as soon as they are obtained, and when equipment to record teaching samples is in limited supply. Extensive in-class observations are not essential to five of the six assessment purposes listed earlier in this paper.

The cost of assessment depends upon both the type of performance to be measured and the assessment procedure to be used. Peer observations are the least expensive; in-class observations the most costly unless large amounts of equipment must be purchased to install critiquing of recorded lessons. Table 3 presented an example of the costs associated with assessment of microteach samples of teaching. Estimates of the cost of trained observer scoring of the same number of teaching skills in the classroom would be higher than the $72 per teacher projected for the microteach sample. Because of this increased cost, in-class observations probably should be used only when no other assessment method will provide the needed information.

Developing the Teachers' Skill

A basic assumption of teacher assessment is that once data are available regarding a teacher's performance, training will be offered to improve that teacher's competence. Using his skill profile to identify areas of strength and weakness, the teacher may waive training in one skill area to emphasize improving another. To be successful, a program that allows these options incorporates training packages that cover a diversity of skills. Within a performance orientation, each of these packages generally will include models of the skills to be acquired, opportunities to practice the skills, and evaluative feedback regarding the teachers performance during practice.

As noted by Joyce (1973), development of such training materials is costly. Time and effort, represented by dollars, must be spent to identify the skills to be included in the package, develop the initial form of the materials, test the materials to be sure teachers acquire the specified skills, and revise the materials based upon testing results. Table 4 presents cost data for the design of one such type of training package, the Minicourses developed at the Far West Laboratory. The data represent average costs for each development task based upon the design and testing of four Minicourses.

The merit of building a training system that uses as many already developed materials as possible is obvious given such development costs.

An additional cost factor that should be considered as part of a complete assessment cycle is the cost of providing training. Table 5 lists the estimated costs for a 5 week training sequence that includes microteaching. Start-up costs such as purchase of training materials

Table 4

Cost of Developing Training Materials

Development Task	Cost
Conceptualization and skill identification	$ 14,670
Product development	28,962
Product testing (one test with 30-50 teachers)	46,279
Revision after testing	12,627
TOTAL	$102,538

and video equipment are included as well as the cost of conducting the training itself. Once the training materials and equipment are available, this type of training runs approximately $100 per teacher. Given a shorter training period, a larger number of teachers to work with the coordinator, or less microteaching, the costs would be less. Ultimately, the cost would depend upon whether the training was part of a larger performance-based assessment and training system or a single, one shot, effort.

Table 5

Cost of Training*

Start up costs		
Materials purchase		
16mm filmed models of teaching skills		$1,000
teacher handbooks 30 @ $2.50		75
videotapes for microteaching 30 1/2 hour tapes		450
Equipment purchase		
VTR equipment, 3 sets		6,000
	Subtotal	$7,525
Training costs		
Coordinator of training, 6 weeks for planning and monitoring		2,025
**Substitutes to release teachers, 1 substitute per 10 teachers x 5 weeks of training = 30 days @ $35/day		1,050
	Subtotal	3,075
	TOTAL	$10,600

* Based upon Minicourse training costs.
** Optional; other ways may be used to release inservice teachers for microteaching; not required in preservice

Summary

In this paper questions have been raised regarding:

- the purposes of assessment;
- the teacher performance to be assessed; and
- the individuals who will do the assessment, particularly the level of training and skill they must have based upon the assessment approach being used.

In each question area, several optional courses of action were presented. Different cost levels were associated with each option. Thus, the task for the individual, or agency, responsible for assessing teacher performance is to determine which options within each decision area best meet the policy-making and training information needs of that particular program (agency). Central to this decision will be the quantity and quality of data required given the purposes for which the findings will be used. A reasonable balance must be achieved between the desire to obtain detailed, highly reliable data and the total dollars to be spent.

Bibliography

Davis, O.L. and Tinsley, D.C. Cognitive objectives revealed by classroom questions asked by social studies teachers. Peabody Journal of Education, 1967, 45, 21-26.

Dodl, Norman, et al. A Catalog of Teacher Competencies--A Working Document. Tallahassee: Florida State University, 1972.

Elementary Teacher Education Models. A Summary, Journal of Research and Development in Education, Vol, 2, No. 3. University of Georgia, Athens, Georgia, Spring 1969.

Gallagher, J.J. Expressive thought by gifted children in the classroom. Elementary English, 1965, 42, 559-568.

Guszak, F.J. Teacher questioning and reading. The Reading Teacher, 1967, 21, 227-234.

Herbert, John. A research base for accreditation of teacher preparation programs. Accreditation and Research Problems, J.L. Burdin and M.T. Reagan, Eds. ERIC Clearinghouse on Teacher Education, Washington, D.C., 1971.

Houston, W. Robert. Strategies and Resources for Developing a Competency-Based Teacher Education Program. University of Houston, New York State Education Department and Multi-State Consortium on Performance-Based Teacher Education, 1972.

Husek, T.R. and Sirotnik, Ken. Matrix sampling in educational research: an empirical investigation. Paper presented at the 1968 convention of the American Educational Research Association, Chicago, Illinois.

Joyce, Bruce R. Estimating Costs of a Competency Orientation. Paper presented at Multi-State Consortium on Performance-Based Teacher Education, New Orleans, February 1973.

Joyce, Bruce R. and Weil, Marsha. Models of Teaching. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Lord, F.M. and Novick, M.R. Statistical Theories of Mental Test Scores. Reading, Mass.: Addison-Wesley, 1968.

Rosenshine, B. Teaching Behaviours and Student Achievement. London: National Foundation for Educational Research in England and Wales, 1971.

Silberman, C.E. Crisis in the Classroom. New York: Random House, Inc., 1970.

Smith, B.O. Certification of educational personnel. University of South Florida, unpublished paper, 1971.

Turner, R.L. Relationships between Teachers for the Real World and the elementary models programmatic themes and mechanisms payoffs, mechanism and cost. In B. Rosner (ed.), The Power of Competency-Based Teacher Education. Boston: Allyn and Bacon, Inc., 1972.